1 Introduction

1.1 Motivation

Propaganda has long been a tool of political influence, but in recent years it has taken a new online form: fake news. Fake news, once a buzzword on the internet, is now at the center of global politics, one of the most common bigram in the lexicon of United States president Donald Trump. After the term gained prominence in Trump's presidential campaign in 2016, it exploded into public conciousness, earning the distinction of Webster-Collins' "Word of the Year" in 2017.[1] As the current presidential race unfolds, fake news has returned to the center of the conversation, with major social media companies facing scrutiny of their misinformation policy. This phenomenon is also not a distinctly American problem–investigate reporting both during and after the 2018 Brazilian president election demonstrated that more than 40% of right-wing viral news articles shared on the popular messaging service WhatsApp were fake news favoring the eventual winner, Jair Bolsonaro.[2]

Fake news is not only politically significant but also dangerously tempting. Studies have shown false content propagates faster through social media than real content.[3] Blatantly false or exaggerated rhetoric can even lead to violent action, as demonstrated by the "Pizzagate" incident in which a man stormed a D.C. pizzeria with an AR-15, having been convinced by false and unverified information that a pedophile ring operated out of the restaurant's basement.[4] Fake news can also be incredibly easy to create. In 2019, a group of researchers at the Allen Institute for Artificial Intelligence published text generation model able to produce fake news.[5] In a troubling conclusion, the researchers found that state-of-the-art fake news detection systems struggled more with identifying fake news produced by their systems than actual fake news.[6] Fake news is thus easy to create, spreads quickly, and is hard to detect, a dangerous combination making it a serious threat to civic society.

[1] [?]
[2] [?]
[3]
[4]
[5]
[6]

1.2 Our Contribution

Given the danger that fake news poses, machine learning and natural language processing researchers have devoted significant attention to the problem of fake news classification. However, previous attempts at fake news classification overwhelmingly rely on highly complex models suffering from the "black box" problem. As a result, these models lack interpretability and do not allow us to reach new conclusions about the nature of fake news. In order to begin closing this gap, this research adopts an interpretable approach, with the overall objective of a producing a fake news classification model with comparable accuracy to state of the art models without compromising interpretability.

1.3 Organization

In Chapter **??** we give…, in Chapter **??**…